NPS-CS-11-006

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

A COMPARATIVE ANALYSIS OF FILE CARVING SOFTWARE

by

Timothy Courrejou
Simson L. Garfinkel

September 12, 2011

Approved for Public Release; Distribution is Unlimited

NAVAL POSTGRADUATE SCHOOL
Monterey, California 93943-5000

This report was prepared for and funded by the Defense Intelligence Agency, Washington, DC.

This report was prepared by:

Timothy Courrejou
Department of Computer Science

Simson L. Garfinkel
Department of Computer Science

Reviewed by:

Peter Denning
Chairman
Department of Computer Science

Released by:

Karl A. van Bibber, Ph.D.
Vice President and Dean of Research

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*
12–9–2011

2. REPORT TYPE
Technical Report

3. DATES COVERED *(From — To)*
2010-06-14—2010-08-20

4. TITLE AND SUBTITLE
A Comparative Analysis of File Carving Software

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Timothy Courrejou, Simson L. Garfinkel

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943

8. PERFORMING ORGANIZATION REPORT NUMBER
NPS-CS-11-006

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DIA, Washington, DC

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution is Unlimited

13. SUPPLEMENTARY NOTES
The views expressed in this report are those of the author and do not necessarily reflect the official policy or position of the Department of Defense or the U.S. Government.

14. ABSTRACT
Though there has been significant research into file carvers, there has been little comparison or validation of different file carvers. Such comparison and validation is vital if the state of the art is to progress. We present a methodology for comparing file carvers based on realistic data and present the results of applying the carver to the Foremost, Scalpel, PhotoRec, and Adroit, carvers.

15. SUBJECT TERMS
Residual Data

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(include area code)*
FOUO	Unclassified	Unclassified	UU	22	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 8–98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

Abstract

Though there has been significant research into file carvers, there has been little comparison or validation of different file carvers. Such comparison and validation is vital if the state of the art is to progress. We present a methodology for comparing file carvers based on realistic data and present the results of applying the carver to the Foremost, Scalpel, PhotoRec, and Adroit, carvers.

1 Executive Summary

We analyzed four carvers that are widely used by computer forensics professionals using realistic data previously created by our research group [1]. We tested the carvers on complete, intact file systems and compared their ability to locate the allocated files resident inside each file system. This is a fair test because carvers should, at very least, be able to find allocated files, and the fact that the files could be recovered gave us an unambiguous "ground truth" to rate our carvers against.

Among our overall findings:

1. None of the carvers were able to find all of the allocated files of a given file type.
2. Scalpel consistently generated the largest number of carving results, but the smallest number of useful carving results. (That is, it had the highest rate of false positives or junk carves.)
3. Photorec was generally the top performer, although Adroit was able to occasionally find files that PhotoRec could not.

This is an interim report; additional work is required to determine if any of the carvers found legitimate files that were not found by the other file carvers. Nevertheless, the results of this report indicate that large scale media processing that requires carving should be done with multiple carvers that have their results combined, and that there is no single "best" carver.

2 Background

"File Carving" in a technique used in both computer forensics and data recovery to construct files based on file contents rather than using file system metadata. In computer forensics processing, file carving is commonly used to recover files from the unallocated space of file systems. Typically carving is used for files that were once allocated within the file system and were later deleted: the file system directory and metadata is no longer available, but some or all of the original files remain. File carving is likewise useful for recovering files when a partition has been reformatted. When entire files cannot be recovered, file carving can recover useful fragments—for example, file carving can sometimes recover the icons from within a JPEG file when the entire file cannot be recover.

File carving is also useful in data recovery when critical file system structures such as directories, file allocation tables, and master file tables are rendered inaccessible due to device failure.

There are many file carvers available for forensic use. The simplest of these carvers locate the headers and footers for known file types on the computer's hard drive and write the header, the intervening bytes, and the footer to a file. The user then attempts to open the file with a conventional application program: if the file can be opened, the file is validated. More sophisticated file carvers perform this validation *before* the file is written to the disk; these are called *validating file carvers*. The most sophisticated file carvers can reassemble fragmented files; these are called *fragmented file recovery*.

3 Test Methodology and Configuration

This section documents our testing methodology and configuration.

3.1 Carving Intact File Systems

This project tested the ability of file carvers to recover files from complete, intact file systems. File carvers are typically not used to recover files from such file systems, as in these cases the file system metadata can be used to recover the file contents. However there is no reason that a carver should not be able to recover files from an intact file system. Indeed, the problem of recovering files from an intact file system is quite similar to the problem of recovering files from a file system that has been formatted with the Windows XP `format` command—a command that wipes some file system metadata but leaves the files largely intact and in their original locations.

The advantage of carving from intact, complete file systems is that the ground truth for the allocated files is known. At very least, a good file carver should be able to recover all of the allocated files. A file carver might also be able to recover deleted files, or partially overwritten files, or files extracted from compound documents. This study did not evaluate the ability of file carvers to extract such data.

Some file carvers perform *fragment recovery carving*— that is, they can recover files that are split between more than one physical location on the hard drive. Although some of the allocated files in this study were fragmented, we did not evaluate which carvers were best at recovering fragmented files.

Some file carvers perform *validation* to avoid recovering data objects that appear to be useful files but are not. For this study we limited our evaluation of the carved data to comparison with allocated files; we did not evaluate carving results that were not allocated files.

3.2 Carver Configuration

Performance of carvers is highly dependent upon the carver's configuration. Because there are many configuration options, we decided to run each file carving program with its default configuration unless the default configuration resulted files of interest not being carved (as was the case with scalpel), or the default configuration caused the carver to not function properly.

3.3 File Types of Interest

Although the evaluation methodology in this report can be used with any file type, this report restricts itself to the following file types:

- bmp
- doc
- gif
- jpg
- png
- ppt
- xls
- zip (including docx, pptx, xlsx, and other zip-based file types)

3.4 Carvers

Carvers were chosen for this experiment based on their availability and use within the forensic community:

Foremost-1.5.7 is an open source carver developed in the 1999s that has been infrequently supported since then. Foremost uses a configuration file containing file type definitions which determine the types of files which Foremost will attempt to carve from a disk image. The default configuration file `foremost.conf` contains definitions for many common file types which included the types of files that were of interest to us.

Scalpel-1.60 is an open source file carver based on Foremost-069, which carves files from a disk image by reading a configuration file containing header and footer definitions for the file types to be carved [2]. For each header bit sequence matching a file type definition in the configuration file that is found, Scalpel copies the sequence of bits from the start of the header to the location on the disk image where the next footer for that file type is encountered, or the allowed file size for the file type is reached.
Scalpel's configuration file `scalpel.conf` contains a list file types that are to be carved; the default configuration file has each line commented out, causing zero file to be carved when using the default configuration file. We un-commented definitions for a large number of files types; the un-commented lines can be found in Table 1, in the appendix.

PhotoRec-6.11.3 by default is configured to carve more than 320 file types. File system information, block (or cluster) sizes and file header, and footer, bit-sequences are used by PhotoRec in order to decrease false positives, and allow for fragmented file recovery. PhotoRec was run without modifying the default configuration.

Adroit Photo Forensics 2010 Table 6 in the appendix shows configuration of Adroit that was used during our experiments. One notable change to the configuration was made. First, the option to generate hash values for carved files was deselected since we would be computing the hash values of all carved files during the analysis.

3.5 Carving Targets

The images carved during this experiment were selected because they were representative of document types likely to be present in cases where file carving is needed. These images were:

nps-2009-canon2-gen6.raw is the last in a set of six FAT32 forensic images created during a typical use of a Canon PowerShot SD800IS digital camera. The images were made by placing an SD card into the camera, taking photos, removing the card, erasing select photos, imaging the card, and then repeating the process. Some of the JPEGs are fragmented, some are not. Some are allocated in the file system, some are deleted (not allocated) but recoverable, and several have data present but no longer have any file system metadata and can only be recovered through carving. Of these carvable JPEGs at least two are fragmented. This image was created to test and teach basic file recovery, fragmented file recovery, and file carving.

nps-2009-ubnist1-gen3.raw is the final image in a set of three made from a USB memory stick that contains a bootable copy of Ubuntu 8.10 Linux. Over the course of several weeks the image was repeatedly booted in Linux, used to browser several US Government websites, and then shut down and imaged. This image contains a boot loader and a FAT32 file system.

nps-2009-domexusers.raw is an NTFS file system of computer running Windows XP containing two user accounts. Over a course of several days, an experimenter playing the role of two users exchanged instant messages and emails with a third user that resided on a separate system. The two accounts received, edited and saved office document files as well as various media files. Some of these files were then deleted. Email and instant messenger conversations were saved locally on the system. The accounts also visited web pages for news and webmail. Details of the precise method by which this disk image were prepared can be found in another publication. This image has been redacted with a special redaction tool that removes the instructions from the Microsoft Windows executables but leaves behind the strings. This allows analysis of the DLLs but prevents the image from being used to run Windows without a license, which believe is sufficient redaction for the purpose of distributing the disk image under the "fair use" provisions of the US Copyright Act.

jo-2009-12-08.raw is one image taken from the "M57 Patents" images, which contains video files.

Each carver was run on each disk image, for a total of 20 carving trials. Due to problems that we encountered with some of the carvers, some of the runs had to be repeated.

3.6 Naming conventions

Each carving run was given a unique name made up of the image that was carved, the carving software that was used, and the number of times that combination had been used before: `IMGNAME-CARVERNAME-carveNUMBER`. Files and directories created during the course of this experiment, were given names derived from the compound name. For each run we produced two items:

- Output from the carver, which was captured in a directory named `IMGNAME-CARVERNAME-carveNUMBER`
- A Digital Forensics XML file which we produced based on the contents of the output directory. The DFXML file was named `IMGNAME-CARVERNAME-carveNUMBER.xml`

4 Test Results

Table 1 shows the results of each carving run with the total amount of data generated by the carver and the amount of time that the carver required. Scalpel's carve directories are all suffixed with `carve2` due to initially carving the disk images without uncommenting the lines, specifying the types of files to carve, in its configuration file.

Table 1: Carve names with their respective output directory sizes, and elapsed time.

Carve Name[a]	Size	Carve time HH:MM:SS.MS
jo-2009-12-08-adroit	68M	00:11:23.
jo-2009-12-08-foremost	2.2G	9:19.36
jo-2009-12-08-photorec	9.9G	5:12.52
jo-2009-12-08-scalpel	461G	2:31:05.
nps-2009-canon2-gen6-adroit	29M	00:00:17.
nps-2009-canon2-gen6-foremost	16M	0:00.73
nps-2009-canon2-gen6-photorec	24M	0:04.50
nps-2009-canon2-gen6-scalpel	2.9M	0:01.36
nps-2009-domexusers-adroit	55M	00:26:23.
nps-2009-domexusers-foremost	3.0G	27:13.32
nps-2009-domexusers-photorec	19G	13:47.95
nps-2009-domexusers-scalpel	461G	2:35:59.
ubnist1.casper-rw.gen3-adroit	19M	00:02:30.
ubnist1.casper-rw.gen3-foremost	116M	0:21.10
ubnist1.casper-rw.gen3-photorec	280M	0:17.12
ubnist1.casper-rw.gen3-scalpel	122G	31:32.91.

[a] The -carve1 or -carve2 notation has been omitted for clarity

5 Carving Analysis

We started processing the carve results by producing Digital Forensics XML [?], files using a Python script that scanned each carve results directory, finding the size and SHA1 of every carved file. The information was stored along with the carved file name in `<fileobject>` elements in the DFXML files. Using the SHA1 hashes of each file, we were able to partition each set of carved files into two subsets of each carve, the set of carved files in the file system that were recovered using `fiwalk`, and the set of unallocated carved files.

5.1 Allocated files carved

Once the carve DFXML files were created, we used another python script that compared the SHA1s found in the `fiwalk` DFXML to the SHA1s of each fileobject in the carve results DFXML file created by our program. The intersection between carved files and the `fiwalk` DFXML indicates the recoverable files that could be successfully carved from the image (see Table 2). Notice that Scalpel carved 0 allocated files from the Canon image; this is the result of an error in the version of the configuration file that is distributed with Scalpel.

Table 2: Number of files carved from each disk image, by each file carver, the files known to be allocated in the file system, and the total number of those allocated files that were carved by each carver.

Carver	Image	Total carved	Total allocated on disk	Total allocated files carved
Adroit	nps-2009-canon2-gen6.raw	37	42	33
Foremost	nps-2009-canon2-gen6.raw	30	42	22
Photorec	nps-2009-canon2-gen6.raw	33	42	29
Scalpel	nps-2009-canon2-gen6.raw	3	42	0
Adroit	ubnist1.casper-rw.gen3.raw	1,163	1,196	114
Foremost	ubnist1.casper-rw.gen3.raw	1,509	1,196	141
Photorec	ubnist1.casper-rw.gen3.raw	4,836	1,196	269
Scalpel	ubnist1.casper-rw.gen3.raw	15,199	1,196	84
Adroit	nps-2009-domexusers.raw	2,343	21,638	2,311
Foremost	nps-2009-domexusers.raw	18,638	21,638	5,011
Photorec	nps-2009-domexusers.raw	24,053	21,638	6,792
Scalpel	nps-2009-domexusers.raw	57,617	21,638	1,111
Adroit	jo-2009-12-08.raw	1,547	24,445	1,517
Foremost	jo-2009-12-08.raw	17,046	24,445	3,948
Photorec	jo-2009-12-08.raw	19,504	24,445	7,958
Scalpel	jo-2009-12-08.raw	128,694	24,445	985

5.2 Allocated files carved by type

Because file carvers find files by an examination of file content, different carvers have different performance on different kinds of files. Table 3 presents the results by file type for each file carver.

In recent years many application developers have adopted the ZIP compression archive as a kind of universal file type. Depending on whether the ZIP file contains compressed XML or compress class archives the ZIP file can be `.docx` file, a `.jar` file, or files for a variety of other types. When ZIP files are recovered, however, the file carver may not identify the ZIP file as being a particular *kind* of zip file.

Thus, when we carved for ZIP files, we actually found files of a variety of types, including:

- `ZIP` - Compressed file archive
- `AMO` - AOL Instant Messenger UI plug-in file
- `WMZ` - Windows Media Compressed Skin File
- `PBZ` - Picasa Button Zipfile
- `JAR` - Java Archive
- `DAT` - OpenOffice.org data file archives
- `OXT` - OpenOffice.org dictionary file archive

PhotoRec was the only carver that successfully carved the allocated ZIP files.

5.3 Validation of carved files by type

Once the set of carved files was found, it was necessary to find the specific carved files that were valid. We validated files using an automated validation script. Image file formats (e.g., png, gif, ppm, bmp, pbm, pgm, jpe, jfif and jpeg) were deemed valid if, using Python's `Image` module could open a validate the image file using this short program:

```
import Image,sys
try:
    img = Image.open(sys.argv[1])
    img.verify()
    print "The image verifies."
    exit(0)
except Exception:
    print "The image does not verify."
    exit(1)
```

Microsoft Office files (e.g. doc, xls, ppt) were deemed to validate if the `wvSummary` tool (part of the wvWare library) could output correct summary information.

Zip files (e.g. zip, docx, xlsx, pptx) were validated by attempting to decompress each of the zip file using `unzip`.

The number of unallocated files, and the percentage of them which passed the validation test, is shown in Table 4. The first two columns list the image from which the files were carved, and the extension, which was used to determine the validation technique. The remaining columns

contain the number of unallocated files carved by each carver per extension per image, and the percent of each set which passed validation.

A surprising percentage of the Foremost files validated. Recall, however, that the files presented in Table 4 were only those files that were allocated in the file system. Although Foremost carved a huge number of files, only a very small number of those files were actually allocated files. Thus, the true positive rates presented in Table 4 is unrealistically high. This large number of successfully validated files may have been caused by carving many subsections of the same file.

A large percentage of the files carved by PhotoRec successfully validated as well, and referring back to Table 1, PhotoRec also had an average carve time that was substantially lower than the other file carvers. This would likely make PhotoRec a good candidate when there is a great need for speed and reliability.

Adroit's results yielded the fewest number of files, the majority of which, passed validation testing. We know from discussions with the Adroit authors that the program is performing its own validation. It is possible that it's validation tests are too stringent.

Considering all of the data presented, we found PhotoRec to perform the best overall. Nevertheless, other carvers found Files that Photorec missed.

6 Conclusion

In this report, analysis performed on file carving data, resulting from carving four disk images with selected file carving programs, has been discussed. We first outlined the data collection procedure which was followed, in which several thousand allocated and unallocated, valid and invalid, files were carved from the set of disk images. Next, we explained our data analysis techniques allowed us to separate the four major groups of carved files which allowed for finding the number of allocated files carved from each disk image and the number of carved, unallocated files which passed an automated file validation testing. Finally, we presented tables containing the analysis results and considered possible conjectures from which could be formulated.

From these results, we determined that PhotoRec was the most effective file carver. This determination was made using the comparative performance of each carver in various categories, including, carving speed, file types and accuracy. Though PhotoRec was thought to be the best choice for a general purpose file carver, situations in which the successful recovery of unallocated image files may benefit from the use of Adroit, which only attempted to recover bmp, jpg, gif and png files, but which claims to be able to recover fragmented files as well. The final two carvers, Scalpel and Foremost, carved a large number of valid files in some file types, though a large portion of invalid files resulted, leaving the laborious job of sifting through invalid carved files with the user of the software.

Based on this work, the overall conclusion is that there is no best file carver—each carver found files that the others did not find. It would seem that the best strategy is to use multiple file

carvers with additional post-processing file validation steps.

6.1 Future work

The analysis presented in this report does not consider the effectiveness of the carvers at recovering fragmented files. It also only considered allocated files; our next report with this same data set shall consider the effectiveness of the carvers at recovering deleted files as well as files that can only be recovered through carving.

We need to re-test the carvers with a procedure that only carves unallocated blocks as identified by the SleuthKit `blkls` command.

Techniques used by file carvers to locate and validate files of various types on a disk image can vary across the numerous file carving utilities which exist to date. Aside from speed and efficiency at which the applications operate, this could also cause carved files to incorrectly validated by automated file validation tools such as the script described in this report. One way to obtain this confidence level in any given automated validation technique is the random sampling of files that both were successfully validated, and which failed validation, manually determining the validity. Comparing the manual validation results with the automated validation results would be a useful indication of the automated program's accuracy.

Another approach to finding the efficacy of file carvers would be to analyze the results of carving a disk image for which there is perfect knowledge of every file which existed on the disk at every instant in time. Gaining perfect knowledge of a disk image could be accomplished by zeroing out a hard drive, then recording all writes to the drive and computing the precise locations of all recoverable objects.

Image	Ext	total	adroit	foremost	photorec	scalpel
jo-2009-12-08	bmp	166	158	154	154	0
jo-2009-12-08	doc	10	0	0	12	4
jo-2009-12-08	gif	1,653	749	857	608	857
jo-2009-12-08	jpg	289	319	312	297	123
jo-2009-12-08	png	362	291	324	320	0
jo-2009-12-08	ppt	4	0	0	0	0
jo-2009-12-08	xls	8	0	0	3	0
jo-2009-12-08	zip	137	0	0	131	0
	Total:	2805	1517	1647	1525	984
nps-2009-canon2-gen6	bmp	0	0	0	0	0
nps-2009-canon2-gen6	doc	0	0	0	0	0
nps-2009-canon2-gen6	gif	0	0	0	0	0
nps-2009-canon2-gen6	jpg	33	33	22	29	0
nps-2009-canon2-gen6	png	0	0	0	0	0
nps-2009-canon2-gen6	ppt	0	0	0	0	0
nps-2009-canon2-gen6	xls	0	0	0	0	0
nps-2009-canon2-gen6	zip	0	0	0	0	0
	Total:	33	33	22	29	0
nps-2009-domexusers	bmp	230	212	147	142	0
nps-2009-domexusers	doc	12	0	0	23	1
nps-2009-domexusers	gif	1,891	778	927	547	929
nps-2009-domexusers	jpg	757	609	384	336	181
nps-2009-domexusers	png	1,353	696	843	827	0
nps-2009-domexusers	ppt	7	0	0	3	0
nps-2009-domexusers	xls	13	0	0	9	0
nps-2009-domexusers	zip	176	0	0	38	0
	Total:	4462	2295	2301	1925	1111
ubnist1.casper-rw.gen3	bmp	0	0	0	0	0
ubnist1.casper-rw.gen3	doc	7	0	0	0	0
ubnist1.casper-rw.gen3	gif	45	0	43	49	43
ubnist1.casper-rw.gen3	jpg	6	97	71	62	41
ubnist1.casper-rw.gen3	png	16	17	27	25	0
ubnist1.casper-rw.gen3	ppt	3	0	0	0	0
ubnist1.casper-rw.gen3	xls	7	0	0	0	0
ubnist1.casper-rw.gen3	zip	0	0	0	0	0
	Total:	95	114	141	136	84

Table 3: Number of allocated files on each image and the number of them which were carved by each carver.

Table 4: Number of valid and invalid, unallocated files by file extension for each carve. The validity of each file was determined by an automated validation script.

Disk image	Ext	Adroit		Foremost		PhotoRec		Scalpel	
		Unalloc. carved files	Valid files	Unalloc. carved files	Valid files	Unalloc. carved files	Valid files	Unalloc. carved files	Valid files
jo-2009-12-08	bmp	2	100.0%	204	37.7%	20	100.0%	983	26.2%
jo-2009-12-08	doc	0	0.0%	8	100.0%	16	50.0%	231	57.1%
jo-2009-12-08	docx	0	0.0%	0	0.0%	0	0.0%	0	0.0%
jo-2009-12-08	gif	0	0.0%	1,154	93.7%	480	99.8%	1,208	92.5%
jo-2009-12-08	jpg	28	3.6%	636	99.7%	147	100.0%	1,186	75.9%
jo-2009-12-08	png	0	0.0%	7,836	99.3%	319	93.7%	10,726	0.0%
jo-2009-12-08	ppt	0	0.0%	0	0.0%	0	0.0%	0	0.0%
jo-2009-12-08	pptx	0	0.0%	0	0.0%	0	0.0%	0	0.0%
jo-2009-12-08	xls	0	0.0%	2	100.0%	2	50.0%	0	0.0%
jo-2009-12-08	xlsx	0	0.0%	0	0.0%	0	0.0%	0	0.0%
jo-2009-12-08	zip	0	0.0%	371	94.1%	7	0.0%	80,248	0.1%
nps-2009-canon2-gen6	bmp	0	0.0%	0	0.0%	0	0.0%	0	0.0%
nps-2009-canon2-gen6	doc	0	0.0%	0	0.0%	0	0.0%	0	0.0%
nps-2009-canon2-gen6	docx	0	0.0%	0	0.0%	0	0.0%	0	0.0%
nps-2009-canon2-gen6	gif	0	0.0%	0	0.0%	0	0.0%	0	0.0%
nps-2009-canon2-gen6	jpg	4	100.0%	8	100.0%	4	100.0%	0	0.0%
nps-2009-canon2-gen6	png	0	0.0%	0	0.0%	0	0.0%	0	0.0%
nps-2009-canon2-gen6	ppt	0	0.0%	0	0.0%	0	0.0%	0	0.0%
nps-2009-canon2-gen6	pptx	0	0.0%	0	0.0%	0	0.0%	0	0.0%
nps-2009-canon2-gen6	xls	0	0.0%	0	0.0%	0	0.0%	0	0.0%
nps-2009-canon2-gen6	xlsx	0	0.0%	0	0.0%	0	0.0%	0	0.0%
nps-2009-canon2-gen6	zip	0	0.0%	0	0.0%	0	0.0%	0	0.0%
nps-2009-domexusers	bmp	0	0.0%	297	95.6%	79	98.7%	888	68.7%
nps-2009-domexusers	doc	0	0.0%	5	100.0%	46	41.3%	550	57.1%
nps-2009-domexusers	docx	0	0.0%	5	40.0%	1	100.0%	0	0.0%
nps-2009-domexusers	gif	26	84.6%	2,752	97.0%	597	99.3%	2,800	96.4%
nps-2009-domexusers	jpg	4	0.0%	813	98.4%	33	100.0%	1,622	69.5%
nps-2009-domexusers	png	2	100.0%	6,989	95.7%	153	88.2%	16,071	0.0%
nps-2009-domexusers	ppt	0	0.0%	2	100.0%	1	100.0%	0	0.0%
nps-2009-domexusers	pptx	0	0.0%	0	0.0%	0	0.0%	0	0.0%
nps-2009-domexusers	xls	0	0.0%	8	87.5%	6	0.0%	0	0.0%
nps-2009-domexusers	xlsx	0	0.0%	4	25.0%	1	0.0%	0	0.0%
nps-2009-domexusers	zip	0	0.0%	128	16.4%	2	0.0%	8,363	0.1%
ubnist1.casper-rw.gen3	bmp	1	100.0%	1	100.0%	1	100.0%	16	6.2%
ubnist1.casper-rw.gen3	doc	0	0.0%	0	0.0%	9	0.0%	37	0.0%
ubnist1.casper-rw.gen3	docx	0	0.0%	0	0.0%	0	0.0%	0	0.0%
ubnist1.casper-rw.gen3	gif	597	77.4%	655	97.9%	426	100.0%	658	97.9%
ubnist1.casper-rw.gen3	jpg	303	95.7%	371	99.7%	217	100.0%	544	73.5%
ubnist1.casper-rw.gen3	png	148	93.2%	211	100.0%	110	96.4%	14	0.0%
ubnist1.casper-rw.gen3	ppt	0	0.0%	0	0.0%	0	0.0%	0	0.0%
ubnist1.casper-rw.gen3	pptx	0	0.0%	4	0.0%	0	0.0%	0	0.0%
ubnist1.casper-rw.gen3	xls	0	0.0%	0	0.0%	2	0.0%	0	0.0%
ubnist1.casper-rw.gen3	xlsx	0	0.0%	0	0.0%	0	0.0%	0	0.0%
ubnist1.casper-rw.gen3	zip	0	0.0%	29	62.1%	0	0.0%	759	0.5%

Appendices

A Hardware Specifications

Hardware	Dell Optiplex 755
Processor	Intel Core^{TM}2 Quad
Internal Hard Drive	
Manufacturer	Western Digital
Model Number	WD2002FYPS-SATA
Capacity	2.0TB
Station 1	
Operating System	Fedora 13 - i386
Dell service tag:	3HHOVG1
Carving software installed:	
PhotoRec	Version 6.11.3
Scalpel	Version 1.60
Foremost	Version 0.69
Station 2	
Operating System	Windows Vista Home Basic 32-bit
Dell service tag:	8HHOVG1
Carving software installed:	
Adroit	Version 1.2
Station 3	
Operating System	Fedora 13 - i386
Dell service tag:	4HHOVG1
Carving software installed:	
PhotoRec	Version 6.11.3
Scalpel	Version 1.60
Foremost	Version 0.69

Table 5: Computer hardware and software configuration used to run file carving experiments discussed report.

B Carver Configuration

```
gif     y     5000000   \x47\x49\x46\x38\x37\x61   \x00\x3b
gif     y     5000000   \x47\x49\x46\x38\x39\x61   \x00\x3b
jpg     y     200000000 \xff\xd8\xff\xe0\x00\x10   \xff\xd9
png     y     20000000  \x50\x4e\x47?              \xff\xfc\xfd\xfe
bmp     y     100000    BM??\x00\x00\x00
avi     y     50000000  RIFF????AVI
mov     y     10000000  ????moov
mov     y     10000000  ????mdat
mov     y     10000000  ????widev
mov     y     10000000  ????skip
mov     y     10000000  ????free
mov     y     10000000  ????idsc
mov     y     10000000  ????pckg
mpg     y     50000000  \x00\x00\x01\xba   \x00\x00\x01\xb9
mpg     y     50000000  \x00\x00\x01\xb3   \x00\x00\x01\xb7
fws     y     4000000   FWS
doc     y     10000000  \xd0\xcf\x11\xe0\xa1\xb1\x1a\xe1\x00\x00
                   \xd0\xcf\x11\xe0\xa1\xb1\x1a\xe1\x00\x00 NEXT
doc     y     10000000  \xd0\xcf\x11\xe0\xa1\xb1
pst     y     500000000 \x21\x42\x4e\xa5\x6f\xb5\xa6
ost     y     500000000 \x21\x42\x44\x4e
dbx     y     10000000  \xcf\xad\x12\xfe\xc5\xfd\x74\x6f
idx     y     10000000  \x4a\x4d\x46\x39
mbx     y     10000000  \x4a\x4d\x46\x36
wpc     y     1000000   ?WPC
pdf     y     5000000   %PDF  %EOF\x0d     REVERSE
pdf     y     5000000   %PDF  %EOF\x0a     REVERSE
mail    y     500000    \x41\x4f\x4c\x56\x4d
wav     y     200000    RFF????WAVE
zip     y     10000000  PK\x03\x04         \x3c\xac
```

Figure 1: Scalpel configuration file. The three character extension in the leftmost column specifies the file types which Scalpel will carve. The 'y' characters in the second column specifies whether the header and footer are case sensitive. The third column contains the maximum carved file size. All subsequent columns specifiy header, footer and carve method information.

Recovery Option	**Selection**
Use file system information found on disk	✔
Show photos not deleted (Active Photos)	✔
Recover using file system objects (LogCarving)	✔
Recover from unallocated space (Normal Carving)	✔
Recover fragmented photos (SmartCarving)	✔
Faster SmartCarving using time limit of (1,200 seconds)	✔
Show thumbnails of all photos recovered	✔
Always create thumbnails from photo dynamically	
Ignore photos smaller than	
Generate MD5 hash of photos	
Generate SHA256 hash of photos	
Generate MD5 hash of evidence	
Generate SHA256 hash of evidence	
Write recovered file information to log	
Photo Formats to Recover	**Selection**
Jpegs (.jpg and .jpeg)	✔
Canon Camera Raw Format (.crw)	✔
Sony Camera Raw Format (.arw)	✔
Windows Bitmap (*.bmp)	✔
Graphics Interchange Format (*.gif)	✔
Nikon Camera Raw format (.nef)	✔
Canon Camera Raw Format (.cr2)	✔
Olympus Camera Raw Format (.orf)	✔
Portable Network Graphics (.png)	✔

Table 6: Adroit file format configuration was accessed via the "Analysis Options" button prior to the start of a carve. Since our analysis included the caluculations of cryptographic hashes for every carved file, the hash calculation options were deselected to speed up the carving duration.

References

[1] Simson L. Garfinkel, Paul Farrell, Vassil Roussev, and George Dinolt. Bringing science to digital forensics with standardized forensic corpora. In *Proceedings of the 9th Annual Digital Forensic Research Workshop (DFRWS)*. Elsevier, Quebec, CA, August 2009.

[2] Golden G. Richard III and V. Roussev. Scalpel: A frugal, high performance file carver. In *Proceedings of the 2005 Digital Forensics Research Workshop*. DFRWS, New York, August 2005. URL `http://www.digitalforensicssolutions.com/Scalpel/`.

Initial Distribution List

1. Research and Sponsored Programs Office, Code 41
   Naval Postgraduate School, Monterey, CA 93943
2. Defense Technical Information Center
   Ft. Belvoir, Virginia
3. Dudly Knox Library
   Naval Postgraduate School
   Monterey, California
4. GDS Program Office, Defense Information Systems Agency
   Fort Huachuca, AZ
   gds@disa.mil